Pergamon

Computers chem. Engng Vol. 19, No. 9, pp. 961–981, 1995

0098-1354(94)00105-7
0098-1354/95 $9.50 + 0.00

NONLINEAR MODEL PREDICTIVE CONTROL OF THE TENNESSEE EASTMAN CHALLENGE PROCESS

N. L. Ricker[1]† and J. H. Lee[2]

[1]Department of Chemical Engineering, BF-10, University of Washington, Seattle, WA 98195, U.S.A.
[2]Department of Chemical Engineering, Auburn University, U.S.A.

(Received 21 March 1994; final revision received 5 August 1994; received for publication 9 September 1994)

Abstract—A nonlinear model predictive control (NMPC) algorithm is developed and tested on the Tennessee Eastman challenge process. The model used in NMPC is a nonlinear, mechanistic, state-variable formulation with 26 states, 10 manipulated variables and 23 outputs. Fifteen unmeasured disturbances and parameters are estimated on-line to eliminate offset in outputs. The NMPC strategy controls eight outputs using eight manipulated variables. There is an additional SISO loop to control reactor liquid level. In direct comparisons, we show that NMPC is always superior to a typical SISO multiloop strategy. There are marked improvements in cases involving large changes in setpoints and/or constraint handling. The NMPC algorithm is similar to previous work, but simplifications in the projection/optimization step make the computations more feasible for a typical process computer.

INTRODUCTION

Several challenge problems for process control were published recently in Volume 17 (1993) of *Computers in Chemical Engineering*. (Most were first presented in a special session at the 1990 Annual AIChE meeting). The authors' intent was to help the control community focus on issues important to industry, and to allow comparison of different approaches on industrially relevant problems.

One such problem is the Tenneessee Eastman "Plant-wide Industrial Process Control Problem" (Downs and Vogel, 1993), which we will hereafter abbreviate as the TE problem. Early work on this problem was reported by Ricker, who developed a simplified dynamic model of the process, and used it to test linear Model Predictive Control (MPC) strategies (Ricker, 1993). Ricker (1995) also analysed the steady-state characteristics of the full Downs and Vogel problem, and published optimal steady-state conditions for eight different operating modes.

McAvoy and Ye (1994) give a detailed study of SISO (multiloop) control strategies. They show that their recommended SISO configuration satisfies the specifications of Downs and Vogel for operation around the base-case conditions. In the present work, we verified their results, and found that performance was indeed good in most cases. As shown later, however, the reactor-pressure control loop causes problems in some situations, especially in transients to new steady-state modes. We also question their decision to control compressor power.

† To whom correspondence should be addressed.

More recently, others have reported similar SISO results (e.g. Banerjee and Arkun, 1993; Desai and Rivera, 1993; Kanadibhotala and Riggs, 1993; Palavajjhala *et al.*, 1993). Palavajjhala *et al.* also considered the use of linear Dynamic Matrix Control (DMC). We review their results in more detail below. In general, however, our exprience suggests that SISO and linear MPC-type algorithms cannot handle the full spectrum of conditions suggested by Downs and Vogel, unless one adds overrides and other logic to cover all the special cases. Such overrides can consume a large part of the design effort, as one must consider many different constraint combinations. Even then, performance can be poor because the SISO strategies are not really designed to deal with multiple, interacting constraints—an important aspect of the TE problem. We were thus motivated to consider other methods.

A natural alternative is to use a nonlinear model in MPC. Early work along these lines has been reviewed by Bequette (1991), and Biegler and Rawlings (1991). The latter give the method the generic acronym NMPC, which we adopt here. In some cases the model is empirical with adjustable parameters (e.g. Morningred *et al.* 1993). We favor the use of a theoretically based model with adjustable parameters. Application studies of this type include Balchen *et al.* (1988), Patwardhan *et al.* (1992), and Wright and Edgar (1994). None has attempted a problem on the scale of the TE process, however.

In the following, we show that NMPC is very effective for the TE problem, providing better results (over a wider range of conditions) than any reported previously. Care must be taken in the modeling and NMPC formulation, however. Modeling and state estimation approaches are discussed in detail by Ricker and Lee (1995), who present a model of the TE process consisting of 26 states, 10 manipulated inputs, 15 disturbance inputs, and 23 outputs. The present work focuses on the NMPC formulation and its application to the TE problem.

We note at the outset that NMPC is not a panacea. For example, a high-pressure override was required for certain extreme conditions. Thus, the use of NMPC did not completely avoid the need for application-specific procedures. The choice of controlled and manipulated variables for NMPC was also non-trivial, and we make no claim that our solution is "optimal". Rather, the intent is to give a balanced viewpoint of the utility of NMPC for such applications.

SUMMARY OF THE TE PROBLEM

The TE problem requires coordination of four unit operations: an exothermic, 2-phase reactor, a flash separator, a compressor, and a reboiled stripper. There are 41 measured output variables and 12 manipulated variables, all of which are indicated on Fig. 1. Measured signals include simulated zero-mean noise. FORTRAN code representing the process is available from Downs and Vogel (1993), but they have chosen not to publish the model equations. Instead, they provide a flowsheet (reproduced here as Fig. 1), a steady-state material balance, and a qualitative description of the key process characteristics.

The main control objective is to maintain measurements of product rate and composition at setpoints, while keeping other variables within specified "shutdown limits". The product consists of chemicals G and H. The base-case production ratio is 50:50 G:H on a mass basis, but Downs and Vogel stipulate operation at 10:90 and 90:10 ratios also. Variables most likely to violate their shutdown limits are the reactor pressure, and the liquid levels in the reactor, separator, and stripper base.

Eleven of the 12 manipulated variables are flow rates, indicated by valve symbols in Fig. 1. Reactants A, C, D and E enter in four separate feeds. Streams 1–3 are essentially pure A, D and E, respectively. Stream 4 is nominally 48.5% A, 51% C, and 0.5% inert B. According to Downs and Vogel, the primary reactions are:

$$\text{Reaction 1: } A(g) + C(g) + D(g) \rightarrow G(l)$$

$$\text{Reaction 2: } A(g) + C(g) + E(g) \rightarrow H(l)$$

There is also formation of an inert (undesired) byproduct, F. The reactions are irreversible and occur in the vapor space of the reactor. Note that each reaction converts 3 mol of gas to 1 mol of liquid. There is a significant holdup of liquid G and H in the reactor, but there is no liquid effluent. To control liquid accumulation in the reactor, one must balance production (by reaction) against vaporization and removal in the gaseous effluent (stream 7).

The reactor effluent goes to a partial condenser that recovers most of the G and H, forming stream 10, which is then stripped to recover volatile reactants. The bottoms from the stripper is the desired product. Byproduct F can leave with the product, which is typically 97.5% G + H. Downs and Vogel stipulate that the product composition is to stay within ±5 mol% of the setpoint.

Uncondensed vapors from the separator recycle to the reactor through a compressor (stream 8). A valve on the compressor recycle allows adjustments in the net flow of stream 8. At the base-case, the molar ratio of stream 8 to stream 11 is about 6:1. There is also a purge to control the accumulation of inerts, B and F. This is the only outlet for B, which is essentially noncondensable (as are reactants A and C).

The remaining manipulated variables are the coolant rates in the reactor and condenser, the stream rate in the stripper, the liquid rates leaving the separator and stripper, and the reactor agitation rate (which is indicated by SC in Fig. 1). See Downs and Vogel (1993) for additional details.

A TYPICAL SISO/MULTILOOP CONTROL STRATEGY

To motivate the NMPC development, we first consider a typical SISO (multiloop) strategy. The key loops† of the McAvoy and Ye (1994) structure are as follows:

1. *Reactor temperature* is controlled by adjusting the setpoint of a slave loop that controls the reactor coolant temperature.
2. *Reactor pressure* is controlled by adjusting the A feed rate (stream 1).
3. *Reactor level* is controlled by adjusting the E feed rate (stream 3).
4. *Production rate* is controlled by adjusting the A + C feed rate (stream 4).

† McAvoy and Ye (1994) designate this as "Scheme 4".

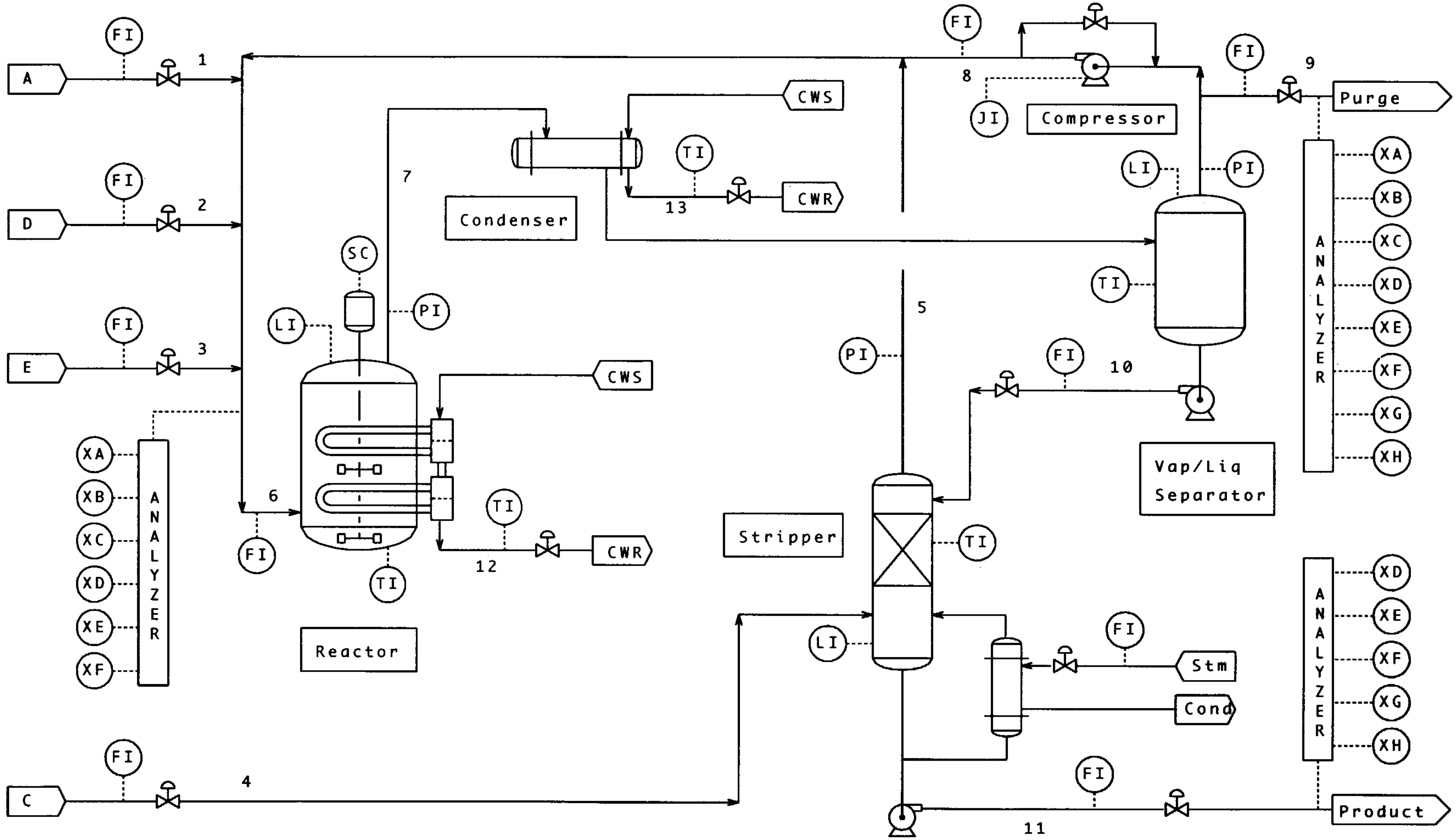

Fig. 1. Flowsheet of the Tennessee Eastman Challenge Process.

5. *Product composition* is controlled by adjusting the D/E ratio (to control the G/H ratio in the product), and the stripper temperature (to control the % of E in the product). The stripper temperature is controlled by a slave loop that adjusts steam rate.
6. *Mol% B in the purge* is controlled by adjusting the purge rate.

Additional loops control the compressor power, recycle flow, and liquid levels in the separator and stripper. Desai and Rivera (1993) have proposed a similar scheme; the main difference appears to be the tuning constants used in the PI controllers, and the manipulation of a D/C rather than a D/E ratio for product composition control. Plavajjhala *et al.* (1993) have tested a similar strategy in comparison to a linear DMC.

We implemented the McAvoy/Ye (1994) scheme using their PI tuning parameters. It often gave good performance; we confirmed the test results given in their paper. On the other hand, disturbances often caused significant upsets in reactor pressure, and there were other shortcomings in special cases. Some of these will be demonstrated in the section on simulation results. Here, we offer the following observations:

1. *Control of reactor pressure*. There are no specifications for variability in reactor pressure, but Ricker (1994) showed that the optimal steady-state condition is as close as possible to the upper shutdown limit of 3000 kPa. If pressure control is poor, one must maintain a relatively large safety margin. The McAvoy/Ye scheme would require a safety margin of at least 100 kPa, equivalent to a 5% increase in operating costs. There are two likely causes of poor performance:
   (a) The A feed may not be available (disturbance 6 of Downs and Vogel, 1993). McAvoy and Ye provided an override for this special case, but it requires that production be reduced by 24%.
   (b) The loop tuning assumes that an *increase* in the A feed rate will *decrease* the pressure. This is valid when the partial pressure of A in the reactor is low. When A is in excess, however, an increase in the A feed rate reduces the partial pressures of the other reactants (and the percentage increase in the partial pressure of A is small). Thus, the reaction rates decrease, gas accumulates, and the pressure *increases*. The magnitude of the loop can vary dramatically. In an extreme case it can change sign (Ricker, 1993).

   An alternative is to control the pressure by adjusting the reactor temperature (Ricker, 1995; Banerjee and Arkun, 1994). Another is to adjust the purge rate.
2. *Control of reactor liquid level*. Ricker (1995) shows that the reactor liquid level should be held near its lower limit at steady-state. Our tests suggest that the McAvoy/Ye scheme provides tight control of this variable. One disadvantage, however, is that any adjustment of the E feed rate causes a proportional adjustment of the D feed rate, which is undesirable (Downs and Vogel, 1993). Thus, rapid setpoint changes would cause problems, but such events should be infrequent. A more important disadvantage is that the E feed is sometimes at its upper limit, in which case a different level-control strategy would be needed. McAvoy and Ye did not provide an override for this case. Alternative manipulated variables include the recycle rate and the separator temperature (Ricker, 1995; Banerjee and Arkun, 1994).
3. *Control of compressor power*. McAvoy and Ye (1994) do not provide a clear rationale for this loop. The power depends on the pressure ratio and flow rate, and is difficult to predict *a priori*. Moreover, the valve on the compressor recycle can only adjust the power over a limited range. If one chooses a setpoint that is either too high or too low for the current conditions, the loop saturates. This is demonstrated later.

We tested several alternative multiloop SISO strategies that were designed to avoid the problems noted above, but there were still cases where one or more loops saturated, leading to poor control. We found the design of overrides for these cases to be difficult and tedious. We also tested MIMO strategies employing linear-time-invariant models (such as DMC, QDMC, IDCOM, etc.), but found that they were too sensitive to the gain variations noted above, and could not be tuned for robust performance. Palavajjhala *et al.* (1993) note that their linear time-invariant DMC works well, but only demonstrate operation near the base case. Although we have not tested their specific DMC design, our experience suggests that serious problems arise when one changes the operating conditions significantly. Moreover, the plots in Palavajjhala *et al.*, show a violation of the ±5% variability specification on product concentration when the production rate decreased 15% under DMC.

Kanadibhotala and Riggs (1993) used Generic Model Control to design the reactor-temperture and stripper-composition control loops, but applied standard PI and ratio control to the remainder of the process. We show below that reactor temperature can be controlled by conventional PI feedback, and the stripper-composition loop is unnecessary. Also, Kanadibhotala and Riggs noted that certain conditions caused stability problems in their reactor-pressure control strategy, which again speaks to the difficulties of pairing SISO loops for this problem. A given pairing will work well for some cases but we have yet to find one that operates equally well for all six steady-state modes and all 20 disturbances.

NONLINEAR MPC (NMPC)

Ricker and Lee (1995) propose a nonlinear (mechanistic), dynamic model of the TE process. A total of 15 parameters are adjusted on line (using an Extended Kalman Filter) to compensate for unmeasured disturbances and model error. Ricker and Lee show that the model captures the essential dynamic behavior of the TE process. It is used here as the basis for NMPC.

The NMPC algorithm used here is a variant of that described by Lee and Ricker (1994). The major modification is that the projection/prediction step uses a *linear* model of the plant, rather than a combination of linear and nonlinear models. This reduces the computational effort significantly (by at least a factor of 3), and has negligible effect on performance in the TE application. The remainder of this section outlines the NMPC algorithm, emphasizing departures from previous work.

Nonlinear process model

The Ricker and Lee's model of the TE process is:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{d}) + \mathbf{w}^x \tag{1}$$

$$\dot{\mathbf{d}} = \mathbf{w}^d \tag{2}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{d}) + \mathbf{v}, \tag{3}$$

where $\mathbf{w}^x$, $\mathbf{w}^d$, and $\mathbf{v}$ are zero-mean, white noise with specified covariances. The $\mathbf{w}^x$ vector represents short-term disturbances having zero mean; $\mathbf{d}$ represents sustained disturbances—integrated white noise. Equation (2) is a well-established way to model disturbances having a step-like character (Garcia *et al.*, 1989). The output vector, $\mathbf{y}$, contains 16 of the 41 measurements available in the TE process. The manipulated variables $\mathbf{u}$, are defined as follows:

$$\mathbf{u}^T = [F_1, F_2, F_3, F_4, F_8, F_9, F_{10}^{\mathrm{p}}, F_{11}, T_{\mathrm{cr}}, T_{\mathrm{cs}}], \tag{4}$$

where F_j is the molar flow of stream j(kmol/h) and T_{cr} and T_{cs} are temperatures in the reactor and the separator, respectively (°C). Stream numbers are as shown in Fig. 1. The state variables, $\mathbf{x}$, are molar holdups of A, B, ..., H at key locations in the process—a total of 26 states (compared to 50 in the Downs and Vogel code).

For the most part, the manipulated variables in the model (equation 4) are the same as those in the plant. The differences are: (1) two manipulated variables in the plant have been omitted from equation (4); (2) two others have been replaced by new variables. The variables omitted are the reactor agitation speed and the stripper steam rate. As discussed later, these are not used in our NMPC strategy. The new variables are T_{cr} and T_{cs}, which replace the two coolant flowrates in the plant. This switch assumes that the corresponding plant tempertures can be controlled by feedback using the coolant rates as manipulated variables; it also eliminates the need for energy balances in the model. Details of these loops are given later.

As shown in Lee and Ricker (1994), linearization and discretization of equations (1)–(3) (for a sampling period of t_s time units) gives the following Extended Kalman filter (EFK):

Model prediction:

$$\begin{bmatrix} \mathbf{x}_{k|k-1} \\ \mathbf{d}_{k|k-1} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{t_s}(\mathbf{x}_{k-1|k-1}, \mathbf{u}_{k-1}, \mathbf{d}_{k-1|k-1}) \\ \mathbf{d}_{k-1|k-1} \end{bmatrix}. \tag{5}$$

Measurement correction:

$$\begin{bmatrix} \mathbf{x}_{k|k} \\ \mathbf{d}_{k|k} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{k|k-1} \\ \mathbf{d}_{k|k-1} \end{bmatrix} + \mathbf{L}_k(\mathbf{y}_k - \mathbf{g}(\mathbf{x}_{k|k-1}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k-1})). \tag{6}$$

Model output estimate:

$$\mathbf{y}_{k|k} = \mathbf{g}(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{d}_{k|k}), \tag{7}$$

where $\mathbf{F}_{t_S}(\mathbf{x}_{k-1|k-1}, \mathbf{u}_{k-1}, \mathbf{d}_{k-1|k-1})$ is obtained by numerical integration of the state equations (1) from given initial conditions, $\mathbf{x}_{k-1|k-1}$, and for constant values of $\mathbf{u}_{k-1}$ and $\mathbf{d}_{k-1|k-1}$, over a single sampling period of duration t_S. $\mathbf{L}_k$ is the EKF gain matrix. Unless noted otherwise, the present work used a time-invariant $\mathbf{L}_k$, obtained as the steady-state solution of the filter equations at the TE base case. Ricker and Lee (1995) show that this simplification causes negligible degradation of estimator performance. Gattu and Zafiriou (1992) also use the steady-state solution, but update it whenever the nonlinear model is re-linearized.

Linearization for prediction/control

Suppose we are at time t_k, and have just obtained estimates $\mathbf{x}_{k|k}$ and $\mathbf{d}_{k|k}$ from equation (6). The current inputs are $\mathbf{u}_{k-1}$, and the problem is to compute $\mathbf{u}_k$, which we will send to the plant (to be implemented from t_k to t_{k+1}). Since $\mathbf{w}^d$ is white with zero mean, our expectation for future values of $\mathbf{d}$ is:

$$\mathbf{d}_{k+j|k} = \mathbf{d}_{k|k}, \quad j \geq 1 \tag{8}$$

i.e. constant. Also, the expectation of $\mathbf{w}^x$ is zero. Thus, we use standard numerical methods† to linearize equations (1) and (2) with respect to $\mathbf{x}$ and $\mathbf{u}$:

$$\dot{\mathbf{x}} \approx \mathbf{f}(\mathbf{x}_{k|k}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k}) + \mathbf{A}_k(\mathbf{x} - \mathbf{x}_{k|k}) + \mathbf{B}_k(\mathbf{u} - \mathbf{u}_{k-1}) \tag{9}$$

$$\mathbf{y} \approx \mathbf{g}(\mathbf{x}_{k|k}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k}) + \mathbf{C}_k(\mathbf{x} - \mathbf{x}_{k|k}) + \mathbf{D}_k(\mathbf{u} - \mathbf{u}_{k-1}), \tag{10}$$

where

$$\mathbf{A}_k = \frac{\partial \mathbf{f}}{\mathbf{x}}\bigg|_{x_{k|k},\, \mathbf{u}_{k-1},\, \mathbf{d}_{k|k}} \qquad \mathbf{B}_k = \frac{\partial \mathbf{f}}{\mathbf{u}}\bigg|_{x_{k|k},\, \mathbf{u}_{k-1},\, \mathbf{d}_{k|k}}$$

$$\mathbf{C}_k = \frac{\partial \mathbf{g}}{\mathbf{x}}\bigg|_{x_{k|k},\, \mathbf{u}_{k-1},\, \mathbf{d}_{k|k}} \tag{11}$$

are matrices of appropriate size.

Feedback stabilization and cascade loops

If one linearizes the model provided by Downs and Vogel (1993) at their base-case steady-state conditions, the resulting $\mathbf{A}_k$ matrix has eigenvalues ranging from -1968 to $+3.07$, i.e. the plant is strictly open-loop unstable. In the absence of feedback, a small perturbation causes a large pressure transient, leading to a shutdown within 1 h. One can use an open-loop unstable model in MPC (e.g., Muske and Rawlings, 1993), but the complexity of the design increases. Moreover, one would like the plant to be stable with NMPC turned off. We thus added stabilizing SISO PI feedback loops to the *plant*. The stabilizing loops are as follows:

1. *Reactor temperature control* using the *reactor coolant valve* as the manipulated variable.
2. *Separator temperature control* using the *condenser coolant valve*.
3. *Reactor liquid level control* using the setpoint of loop 2, above (cascade structure).
4. *Separator liquid level control* using the *separator underflow valve*.
5. *Stripper bottoms liquid level control* using the *stripper underflow valve*.

We have thus assigned four of the plant's manipulated variables, but there is no net change in degrees of freedom because the setpoints of the new loops can be manipulated.

Loops 1 and 2 employ continuous-time PI controllers. Tuning parameters appear in Table 1. A side-benefit is that temperature-related disturbances (3–5, 9–12, 14, and 15 in Table 8 of Downs and Vogel, 1993) are rejected by these loops and have essentially no effect on NMPC performance. Palavajjhala *et al.* (1993) also use lower-level loops in combination with DMC, with the same rationale. Loops 3–5 were simulated as discrete-time PI controllers with sampling periods of 0.01 h. The velocity form given by Seborg *et al.* (1989) was used; Table 1 gives the tuning parameters.

This set of five controllers stabilizes the Downs and Vogel plant, and provides adequate control for many scenarios. For example, it performs as well as the McAvoy/Ye strategy for random (zero-mean) disturbances in the feed composition [IDV(8) $\neq$ 0], but is less effective for a sustained composition disturbance [IDV(1) $\neq$ 0]. The latter requires a sustained change in one or more feed and/or purge rates.

We now return to the stabilization of the linearized version of the Ricker and Lee model. Recall that this model includes loops 1–2 implicitly (through the manipulated variables T_{cr} and T_{cs}; see equation 4). Its linearization gives an $\mathbf{A}_k$ matrix with three near-zero eigenvalues, i.e. essentially an integrating character. In our first attempt to make it strictly stable, we embedded level-control loops 3–5 in the nonlinear Ricker and Lee model prior to linearization. Unfortunately, the resulting linearized

† The MATLAB/SIMULINK routine LINMOD was used for linearization.

Table 1. Tuning constants for SISO loops used for cascade and stabilization

Loop	Controlled variable	Manipulated variable	Controller gain for plant	Controller gain for model	Integral time (τ_I, min)
1	Reactor temp (°C)	Reactor coolant	−8.0%/°C	N.A.	7.5
2	Separ. temp (°C)	Cond. coolant	−4.0%/°C	N.A.	15.0
3	Reactor liq. (%)	Loop 2 setpoint	0.8°C/%	0.8°C/%	60.0
4	Separ. liq. (%)	Separ. underflow	−0.2%/%	−1.36 kmol/h·%	60.0
5	Strip. liq. (%)	Strip. underflow	−0.2%/%	−0.91 kmol/h·%	60.0
6	A feed (kscmh)	Feed 1 valve	0.03%/kscmh	N.A.	0.001
7	D feed (kg/h)	Feed 2 valve	3×10^{-6}/kg/h	N.A.	0.001
8	E feed (kg/h)	Feed 3 valve	2×10^{-6}%/kg/h	N.A.	0.001
9	A&C feed (kscmh)	Feed 4 valve	0.0013%/kscmh	N.A.	0.001
10	Purge flow (kscmh)	Purge valve	0.015%/kscmh	N.A.	0.001

"Plant" gains were used with the simulation code of Downs and Vogel (1993). "Model" gains were used to stabilize the linearized Ricker and Lee nonlinear model.

version exhibited large steady-state offsets in the controlled levels. In other words, the linearization step changed the characteristics of the embedded PI loops.

We, therefore, appended loops 3–5 to the *linearized* model. This required an algebraic (feedback) connection of two linear, continuous-time, state-space systems—the linear model given by the above $\mathbf{A}_k$, $\mathbf{B}_k$, $\mathbf{C}_k$, $\mathbf{D}_k$ matrices, and the system of three independent PI controllers. The inputs F_7, F_8, and T_{cs} are replaced by setpoints for the separator, stripper, and reactor levels, respectively. For loops 4–5, the gains used with the Downs and Vogel model were different from those used with the Ricker and Lee model because the manipulated variables have different units—the former are valve positions (%), whereas the latter are molar flows (kmol/h). In these cases, both controller gains appear in Table 1.

We also defined lower-level flow controllers for the four feeds and the purge. As will be discussed later, NMPC uses the molar rates of these streams as manipulated variables. The flow controllers provide the desired rates, so NMPC need not be concerned with the relationship between valve position and flow rate. Tuning parameters appear in Table 1.† A side-benefit is that these loops reject disturbance 7 (change in header pressure for stream 4). A more subtle advantage is that the loss of stream 1 (disturbance 6) becomes obvious—valve #3 opens to 100%, but there is no flow. This case is discussed in more detail later.

Modification to eliminate direct input/output transfer

The $\mathbf{D}_k$ matrix of the stabilized linear system contains non-zero elements. In particular, the reactor pressure responds instantly to a change in the reactor temperature input (because the temperature affects the vapor pressures of each component), and the production rate responds instantly to a change in input F_{11} (because they are equal, by definition). The usual MPC formulations assume that $\mathbf{D}_k$ is zero for all controlled outputs (e.g. Ricker, 1991). We, therefore, filtered these two outputs through independent, linear, first-order filters with unity gain and time constants of 1/300 h, i.e. very fast relative to the dominant process time constants. When these filters are combined with the above linear model, the number of states increases by two, and the sizes (and certain numerical elements) of the $\mathbf{A}_k$, $\mathbf{B}_k$, and $\mathbf{C}_k$ matrices change accordingly. In the development below, we assume that the linear model has been modified in this manner.

† The model of Ricker *et al.* (1995) did not need such loops because its inputs are molar flows.

It would have been more elegant to modify the nonlinear model to eliminate direct input–output connections (by adding valve lags, for example). On the other hand, arbitrary addition of lags on each input or output would have increased the number of states unnecessarily, and at the beginning of the project, the outputs to be used as controlled variables in MPC had yet to be chosen.

Discretization

We now convert the linearized model (modified as described in the two previous sections) to a discrete-time form. A complication is that the reference point for linearization is the current state, i.e. an *un*steady-state. In other words, the first term on the right-hand side of equation (9) is usually non-zero. (See additional discussion in the next section.)

We, therefore, define a vector of initial conditions:

$$\mathbf{f}_{0k} = \begin{bmatrix} \mathbf{f}(\mathbf{x}_{k|k}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k}) \\ \mathbf{e}_k \\ \mathbf{0} \end{bmatrix}, \tag{12}$$

where $\mathbf{f}(\mathbf{x}_{k|k}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k})$ gives the current time-derivatives of the model states, $\mathbf{e}_k$ is a vector of length 3 containing the current feedback errors (setpoint—measured variable) for the three level controllers, and $\mathbf{0}$ is a zero vector of length 2. The latter gives the initial conditions for the two output filters, which are assumed to be at steady state (due to their speed). The $\mathbf{e}_k$ vector gives the initial conditions of the PI controllers, i.e. the rate-of-change of the integrated feedback errors.

A convenient way to include the effect of initial conditions is to treat $\mathbf{f}_{0k}$ as a column of the $\mathbf{B}$ matrix corresponding to a constant input. In other words, we write equation (9) as

$$\dot{\mathbf{x}} = \mathbf{A}_k(\mathbf{x} - \mathbf{x}_{k|k}) + [\mathbf{B}_k \quad \mathbf{f}_{0k}] \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k-1} \\ 1 \end{bmatrix} \tag{13a}$$

or

$$\dot{\mathbf{x}}^* = \mathbf{A}_k \mathbf{x}^* + [\mathbf{B}_k \quad \mathbf{f}_{0k}] \begin{bmatrix} \mathbf{u}^* \\ 1 \end{bmatrix}, \tag{13b}$$

where $\mathbf{x}$ now includes the states introduced by feedback stabilization and output filtering, $\mathbf{A}_k$ and $\mathbf{B}_k$ are the corresponding linearized model at time t_k, and $\mathbf{x}^* = \mathbf{x} - \mathbf{x}_{k|k}$, $\mathbf{u}^* = \mathbf{u} - \mathbf{u}_{k-1}$ are deviation variables. We then discretize equations (13b) and (10) using standard methods‡ to give:

$$\mathbf{x}^*_{k+j+1} = \mathbf{\Phi}_k \mathbf{x}^*_{k+j} + \Gamma_k \mathbf{u}^*_{k+j} + \Gamma_{0k} \tag{14}$$

$$\mathbf{y}^*_{k+j} = \mathbf{C}_k \mathbf{x}^*_{k+j}, \tag{15}$$

‡ This is described in many texts, and implemented in numerical tools such as C2D in MATLAB.

where for $j \geqslant 0$:

$$\mathbf{x}^*_{k+j} = \mathbf{x}_{k+j|k} - \mathbf{x}_{k|k}, \quad \mathbf{y}^*_{k+j} = \mathbf{y}_{k+j|k} - \mathbf{g}(\mathbf{x}_{k|k}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k}),$$
$$\mathbf{u}^*_{k+j} = \mathbf{u}_{k+j|k} - \mathbf{u}_{k-1}. \tag{16}$$

Linear prediction of future outputs

We next use the above equations to develop a linear prediction of future outputs, which can be used to compute the control actions as in the usual linear MPC formulations. In a manner analogous to that of Lee and Ricker (1994), for a "prediction horizon" of P sampling periods ($P \geqslant 1$) we obtain:

$$\mathbf{Y}_{k+1|k} = \mathbf{Y}^0_{k+1|k} + \mathbf{S}^u_k \Delta \mathbf{U}_k, \tag{17}$$

where

$$\mathbf{Y}_{k+1|k} = \begin{bmatrix} \mathbf{y}_{k+1|k} \\ \mathbf{y}_{k+2|k} \\ \vdots \\ \mathbf{y}_{k+P|k} \end{bmatrix} \quad \Delta \mathbf{U} = \begin{bmatrix} \Delta \mathbf{u}_k = \mathbf{u}_k - \mathbf{u}_{k-1} \\ \Delta \mathbf{u}_{k+1} = \mathbf{u}_{k+1|k} - \mathbf{u}_{k|k} \\ \vdots \\ \Delta \mathbf{u}_{k+p-1} \end{bmatrix} \tag{18}$$

$$\mathbf{Y}^0_{k+1|k} = \begin{bmatrix} \mathbf{C}_k \Gamma_{0k} + \mathbf{g}(\mathbf{x}_{k|k}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k}) \\ \mathbf{C}_k(\Phi_k + \mathbf{I})\Gamma_{0k} + \mathbf{g}(\mathbf{x}_{k|k}, \mathbf{u}_{k+1}, \mathbf{d}_{k|k}) \\ \vdots \\ \mathbf{C}_k \sum_{i=1}^{P} \Phi_k^{P-i} \Gamma_{0k} + \mathbf{g}(\mathbf{x}_{k|k}, \mathbf{u}_{k-1}, \mathbf{d}_{k|k}) \end{bmatrix} \tag{19}$$

and $\mathbf{S}^u_k$ is the matrix of step response coefficients for the linear system, which may be calculated as described by many authors (e.g. Ricker, 1991; Gattu and Zafiriou, 1992).

The vector $\mathbf{Y}^0_{k+1|k}$ gives the output trajectories if no control action were taken. In contrast to the formulations of Lee and Ricker (1994) and Gattu and Zafirious (1992), its calculation avoids a P-step, numerical integration of the state equations. This is especially advantageous when there are many states (e.g. $26+5=31$ in the present case) and/or P is large. In such cases, the P-step integration can be very time consuming—dominating the computational effort. (Note, however, that such integration accounts for the effect of unsteady initial conditions, eliminating the special procedure described in the previous section.)

We originally expected that a P-step integration would provide better performance when there were strong nonlinearities, but the simpler approach worked equally well in our tests. A possible explanation is that the linear approximation for the effect of $\Delta \mathbf{u}$ (which is common to all the algorithms cited above) is limiting the accuracy of the prediction. In that case, improvements in the accuracy of $\mathbf{Y}^0_{k+1|k}$ would not improve overall accuracy significantly.

An altnerative would be to use a nonlinear predictor, as reviewed by Biegler and Rawlings (1991). This introduces daunting numerical problems, however. As will be demonstrated below, performance based on the linear predictor is excellent for the TE problem; a nonlinear predictor might be advantageous in other applications.

The linear predictor used here is very similar to that proposed by Brengel and Seider (1989). It is, if anything, simpler, as the TE problem did not require multiple linearizations of the equations within a sampling period.

Calculation of control action

The control action is the solution of the following (well-established) quadratic program (e.g. Garcia and Morshedi, 1984):

$$\min_{\Delta U} \{\|\Lambda^y(\mathbf{R}_{k+1|k} - \mathbf{Y}_{k+1|k})\|_2^2 + \|\Lambda^u \Delta \mathbf{U}\|_2^2\} \tag{20}$$

subject to

$$\mathbf{u}^{\text{low}}_{k+j} \leqslant \mathbf{u}_{k+j} \leqslant \mathbf{u}^{\text{high}}_{k+j} \quad j = 0, P-1 \tag{21}$$

$$|\Delta \mathbf{u}_{k+j}| \leqslant \Delta \mathbf{u}^{\text{high}}_{k+j} \quad j = 0, P-1 \tag{22}$$

$$\mathbf{y}^{\text{low}}_{k+j|k} \leqslant \mathbf{y}_{k+j|k} \leqslant \mathbf{y}^{\text{high}}_{k+j|k}, \quad j = 1, \mathrm{P} \tag{23}$$

where $\mathbf{R}_{k+1|k}$ is the vector of future setpoints (corresponding to $\mathbf{Y}_{k+1:k}$), and Λ^y and Λ^u are tuning parameters. Future setpoints are assumed known.

Although Rawlings and Muske (1993) have pointed out the advantages of an infinite-horizon formulation of MPC, in the present work we had no difficulties using "blocking" in combination with a finite horizon to achieve nominal stability. As explained by Ricker (1985) and others, blocking is equivalent to setting $\Delta \mathbf{u}^{\text{high}}_{k+j} = 0$ in equation (22) for a specified subset of $j = 0, P-1$. Then only m "moves" (i.e. $\Delta \mathbf{u}$ values) are actually computed, where $m \leqslant P$. The use of $m \ll P$ usually provides nominal stability, and reduces the computational effort.

We emphasize that NMPC controls the *predicted* values of the outputs (see equations 19 and 20). Thus, the model and state estimator are an essential aspect of the strategy. The estimator replaces the "implicit estimator" in a standard MPC (see, e.g., Lee *et al.*, 1994). It must compensate for the measurement delays in the composition analyzers, and eliminate bias (i.e. offset) in the NMPC controlled outputs. As described by Ricker and Lee (1995), this is done by on-line estimation of 15 parameters that adjust reaction rates, vapor-liquid equilibria, and feed compositions in the nonlinear model.

Additional procedural details are available as a MATLAB code.† All simulations were run in

† Send e-mail (ricker@cheme.washington.edu) to N. L. Ricker to request this code. It employes the MPC Toolbox (Morari and Ricker, 1994).

MATLAB. A typical case (48 h of simulated operation) used about 3 h on a Macintosh IIfx. About half this time was consumed in the NMPC calculations, and the other half in the Downs and Vogel TE process simulation.

NMPC FORMULATION FOR THE TE PROBLEM

Assignment of degrees of freedom

We now consider the choice of outputs and manipulated variables for use in NMPC. The problem is related to the pairing of SISO loops in a multiloop (decentralized) strategy. Multiloop analysis for the TE problem has been presented by McAvoy and Ye (1994), Banerjee and Arkun (1993), and Desai and Rivera (1993). Recall that there are 12 degrees of freedom (i.e. manipulated variables) in the plant (see Fig. 1).

Analytical tools are available to support SISO loop pairing, but are of little use for NMPC. This step proved to be one of the most critical and the "best" choice was not obvious at the beginning. The following is an outline of the steps we followed.

A control system for a continuous process must be able to achieve certain steady-state objectives. Previous work on the steady-state aspects (Ricker, 1995) led to the following decisions for NMPC:

- The steam, compressor recycle valve, and agitation rate adjustments were not used; their values were normally held constant during a simulation.†
- The reactor liquid level was controlled by a SISO feedback loop that adjusted the separator temperature setpoint (see discussion of stabilizing loops). This simplified the NMPC design without causing significant performance degradation. The rationale for this SISO pairing is that an increase in the separator temperature causes less liquid to be removed from the recirculating gas. Thus, liquid accumulates in the reactor.

  In simulations the reactor liquid-level setpoint could be varied (outside the NMPC). It was normally held at a low value (such as 60%), which minimizes operating costs as discussed by Ricker (1995).
- Control of the reactor pressure was assigned to NMPC because the upper limit on pressure could constrain the production rate. Also, many of the adjustments affect pressure; MIMO compensation for interactions should improve performance relative to the McAvoy/Ye strategy.
- The rate and %G in the product (stream 11) are specified quantities and must be controlled. We do not attempt to control the (trace) amounts of D, E, and F in the product, as this would require manipulation of the steam rate. As will be demonstrated below, these do not vary significantly. Thus, there is only one independent composition variable in the product—e.g., %G, %H, or their ratio. We have chosen to control %G. The use of %H would be equivalent.

  McAvoy and Ye (1994) control the G/H ratio in the product. We prefer composition because the ratio is sensitive to small variations as the denominator goes to zero, and insensitive in the other extreme. To compensate for this one would have to vary the corresponding Λ^y term in the NMPC objective. Also, specifications on product variability are in terms of mol%, not the ratio.
- The levels in the separator and stripper must be kept within specified bounds.
- The %B in the purge is a logical controlled output. Downs and Vogel (1993) include a case in which this variable must be held at a setpoint. From a more fundamental point of view, the purge is the only outlet for B; it must be manipulated to prevent an uncontrolled accumulation of B, which would decrease reaction rates. McAvoy and Ye (1994) also control the %B. An interesting alternative is the control of %F in the purge, as suggested by Banerjee and Arkun (1993). We did not test this idea; it would be difficult to specify a setpoint for the %F, and its accumulation is less likely, as it also exits with the product.

To summarize, of the 12 degrees of freedom in the plant, the following four are not used by NMPC: compressor recycle valve, condenser coolant, agitation rate, steam rate. The eight remaining are the four feed rates, the purge rate, and the setpoints for the reactor temperature, separator liquid level, and stripper liquid level loops; all eight are used by NMPC.

In the above discussion, we have identified six outputs that must be controlled by NMPC. We thus have $8-6=2$ unassigned degrees of freedom. We could leave these unassigned; MPC does not require the controlled system to be "square", i.e. to have an equal number of controlled and manipulated variables. In the TE problem, however, at a given reactor temperature and vapor volume, a range of reactant partial pressures will yield the desired reaction rates (see rate laws in Ricker and Lee, 1995). If the system has unassigned degrees of freedom, these partial pressures may drift into a region that is more costly (e.g. increased purge of a valuable chemical) or vulnerable to certain disturbances.

† Occasionally they were changed step-wise to achieve a particular end. Such exceptions are noted in the text.

Table 2. Manipulated and controlled variables used in NMPC

No.	Manipulated variables Description	Λ_i^u	No.	Controlled variables Description	Λ_i^y
1	Feed 1 rate, F_1 (kmol/h)	0.2	1	Production rate, F_{11} (kmol/h)	0.6
2	Feed 2 rate, F_2 (kmol/h)	0.2	2	Reactor pressure, P_r (kPa)	0.034
3	Feed 3 rate, F_3 (kmol/h)	0.1	3	A in react. feed, $100y_{A6}$ (mol%)	0.2
4	Feed 4 rate, F_4 (kmol/h)	0.2	4	E in react. feed, $100y_{E6}$ (mol%)	0.2
5	Purge rate, F_9 (kmol/h)	0.3	5	B in purge, $100y_{B9}$ (mol%)	0.2
6	Reactor temp., T_{cr} (°C)	0.4	6	G in product, $100x_{G11}$ (mol%)	2.0
7	Sep. liq. level setpoint (%)	0.02	7	Sep. liq. level (%)	0.05
8	Strip. liq. level setpoint (%)	0.02	8	Stripper liq. level (%)	0.05

One way to avoid this is to define targets for a subset of the manipulated variables. Grosdidier *et al.* (1988) have included this feature in the IDCOM-M controller—the "ideal resting value". In our formulation, one could define one or more inputs as outputs (with setpoints), which would be equivalent. None of the remaining manipulated variables have obvious targets, however. We tested a target for the reactor temperature, but a poor choice caused the process to operate non-optimally. We also considered targets for one or more of the feed rates. The rates of streams 1 and/or 4 would be a poor choice, as these must respond to sustained disturbances in the concentration of stream 4. Thus, a target on one of these would cause steady-state offset in the other (more important) controlled variables. Similarly, a target on the flow of D or E would, in general, be impossible to satisfy at steady-state while meeting the setpoints on product rate and composition.

We, therefore, defined setpoints for two reactant concentrations. The "optimal" concentration of A in the feed to the reactor is nearly constant over a wide range of production rates and compositions (Ricker, 1995). Thus, the mol% of A in the feed was selected as a controlled output. We chose the %E in the reactor feed as the other controlled variable. The %C might have been a better choice, but was not tested in the present work. A disadvantage of the %E is that the optimal concentration of this limiting reactant depends strongly on the desired %G in the product (Ricker, 1995).

Table 2 summarizes the eight controlled and eight manipulated variables selected for NMPC. It also shows the corresponding symbols in the model of Ricker and Lee (1995), the measurement units, and the nominal penalty weights (elements of Λ^y and Λ^u in equation 20). Note that all eight NMPC manipulated variables are setpoints of lower-level feedback loops (defined in Table 1).

Penalty weights, prediction horizon, and blocking

Magnitudes of the penalty weights reflect the inherent scales of the variables. For example, a 20 kPa variation in reactor pressure is much less serious than a 20 mol% variation in the product concentration, hence the relatively large penalty on the latter. Other considerations in the choice of weights are as follows:

- Product rate is to be kept near its setpoint.
- Tight control of A, E, and B concentrations is unnecessary. Their penalties are an order of magnitude less then the penalty on %G in the product.
- Levels in the separator and stripper should be allowed to vary (over the short-term) in order to hold production rate, hence these penalties are small.
- Large changes in F_1, F_2, F_4, and T_{cr} are undesirable, and penalties in Λ^u reflect this. Variations in other manipulated variables are less critical.

The main problem in selecting the penalty weights is balancing—absolute values are unimportant. We quickly found a set that gave reasonable results, and it appeared that large variations would give equally good performance. This is consistent with experience on other MPC applications (e.g. Gelormino and Ricker, 1994). Similarly, the choice of the prediction horizon, P, and blocking factors was non-critical. The results reported here used $P = 10$, and blocking factors of $M = [2, 3, 5]$, i.e. three moves of the manipulated variables.

The sampling period for NMPC was 0.1 h, the same as that for the state estimator (Ricker and Lee, 1995), and the gas sampling system. Thus, the prediction horizon was equivalent to 1 h of operation, which, incidentally, is much less than the response time of some of the controlled variables (the %B in the purge, for example). On the other hand, other controlled variables, such as the pressure, respond very rapidly. Thus, the chosen sampling period is a compromise—a situation that one encounters in most industrial MPC applications.

Constraints

We now consider constraint specifications, another critical decision. Table 3 gives the *nominal* constraints for the manipulated variables. Note that these variables are setpoints for lower-level flow, temperature, or level controllers. As discussed earlier, this frees NMPC from control of fast-acting variables. On the other hand, it becomes difficult to define the *true* constraints on the NMPC manipulated variables. For example, the maximum purge rate, F_9, depends on the pressure drop across the purge valve, which can vary. In a real process, it could also depend on valve wear, extent of dirt buildup, etc.

Our approach included on-line adjustments of the constraints for the flows (first 5 variables in Table 3). If a particular valve was wide open but its measured flow was below the expected maximum value, we changed the constraint to be slightly higher than the measured value. If the measured rate later reached or exceeded this maximum, we increased the constraint by a prechosen amount (but never exceeded the nominal values in Table 3). This simple strategy worked well for the TE problem, but might fail in other applications. An alternative would be to model the valve characteristics, but then one is vulnerable to modeling error. The coordination of MPC with lower-level loops is an important practical problem that deserves further research.

The adjustments in the upper bounds on flows are critical for one of the disturbances—the loss of the A feed. In that case, the rate of stream 1 is zero, regardless of the valve setting. If this were not communicated to NMPC, it would continue to request (and expect) A to be fed, leading to poor performance. Our strategy works well in this difficult case, as will be demonstrated later.

The constraints on the reactor temperature were chosen to keep it within "reasonable" bounds. If T_{cr} goes much higher than 128°C, the rate of byproduct formation becomes excessive. Our bound is far below the shutdown limit (175°C).

Constraints on the liquid level setpoints were chosen to keep the levels within the shutdown limits. The (tightly tuned) level-control loops are actually responsible for keeping the levels within their bounds. The valves on streams 7 and 8 provide a large flow range, so these loops never saturate.

An alternative would be to eliminate the PI level controllers and allow NMPC to use F_7 and F_8 as its manipulated variables. However, we would then need constraints on the controlled levels, which in this case would be *outputs*. As discussed by many authors (e.g. Zafiriou and Marchal, 1991; Rawlings and Muske, 1993), the use of output constraints is theoretically possible, but can make tuning more difficult. We experimented with output constraint formulations in the present work and concluded that they were undesirable. Our use of cascaded level loops (with constraints on their setpoints), combined with small penalties on the measured levels in NMPC, represents a very effective "averaging level control" strategy (see McDonald *et al.*, 1986, and Ye and McAvoy, 1994, for alternative approaches). As will be demonstrated later, variations in product rate are eliminated under most conditions.

The remaining critical constraint is the upper shutdown limit on the reactor pressure, P_r. We usually specify a setpoint of 2850 kPa (or less) for P_r, and the system easily keeps it within 50 kPa of this value. There are certain disturbances that cause larger excursions in pressure, however. The loss of the A feed is the most difficult to handle.

Again, we experimented with both "hard" and "soft" constraints on P_r in NMPC, but could not achieve satisfactory results. The fundamental problem is a conflict between the P_r constraint and the other control objectives, which cannot be resolved by tuning the penalty weights. When the A feed (stream 1) shuts down, it is impossible to provide the 1:1 stoichiometric ratio of A and C. The stream 4 rate must increase (to provide enough A to maintain the reaction rates). Since the concentration of C in stream 4 is greater than that of A, the purge rate must increase to eliminate the excess C. When the production rate setpoint is high, however (as in the base case), the purge valve saturates, and the pressure begins to increase rapidly. At this point, "something has to give".

When we used a hard constraint on P_r, NMPC made drastic moves in other variables in an attempt to satisfy the constraint. These moves were often in unexpected (and unacceptable) directions. With a soft constraint (e.g. Ricker *et al.*, 1989; Zafiriou and Chiou, 1993), one must tune its penalty weight to avoid a shutdown without causing large offsets in the other controlled variables; we could not achieve an acceptable balance.

A more robust strategy is to force NMCP to degrade in a desired direction. In the TE problem, excessive pressure indicates that the current production rate setpoint is infeasible. We, therefore, include an override loop that reduces the production rate *setpoint* whenever P_r goes above a specified limit; $P_{r,\max} = 2920$ kPa was used for all simulations

Table 3. Nominal constraints on manipulated variables

No.	Variable	Min.	Max.	Δ Max.
1	Feed 1 rate, F_1 (kmol/h)	0.1	45	10
2	Feed 2 rate, F_2 (kmol/h)	1.0	181	10
3	Feed 3 rate, F_3 (kmol/h)	1.0	181	10
4	Feed 4 rate, F_4 (kmol/h)	1.0	681	20
5	Purge rate, F_9 (kmol/h)	0.1	32	10
6	Reactor temp., T_{cr} (°C)	115.0	128	2
7	Sep. liq. level setpoint (%)	10.0	90	10
8	Strip. liq. level setpoint (%)	10.0	90	10

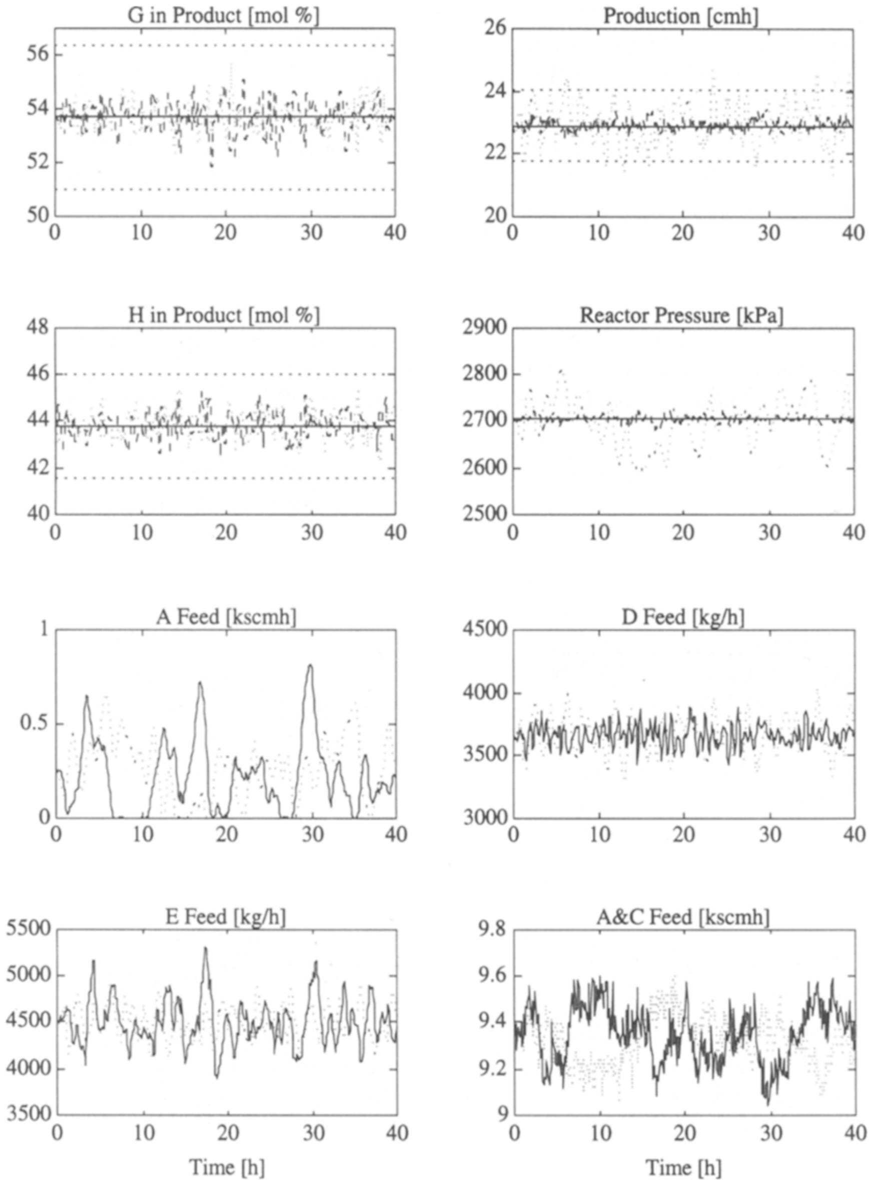

Fig. 2. Comparison of NMPC (solid lines) and SISO control (dotted lines) for IDV(8)—random disturbances in A, B, and C concentrations in stream 4.

reported here. Ricker (1993) describes a similar approach for a simplified version of the TE problem.

The override loop is a PI controller† that attempts to hold $P_r = P_{r,\max}$. Its output (*which is constrained to be non-negative*) is subtracted from the nominal production rate setpoint. Thus, when $P_r > P_{r,\max}$, the override controller's output is positive, and the production rate setpoint decreases. When $P_r > P_{r,\max}$, the override output goes to zero (its lower bound), and the production rate setpoint returns to its nominal value. This strategy *maximizes* the production rate at the given value of $P_{r,\max}$. We demonstrate its performance in the next section.

† The velocity form of the discrete PI controller was used with a sampling period of 0.1 h, a gain of −0.3 kmol/h per kPa, and an integral time constant of 20 min.

NMPC SIMULATION RESULTS

Feed composition disturbances

For most of the load disturbances proposed by Downs and Vogel (1993), NMPC performance is comparable to (but better than) the results published by McAvoy and Ye (1994). For example, Figure 2 shows the results for random variations in the A, B, and C composition in stream 4 [IDV(8)≠0], which is one of the more difficult cases.‡ Both methods keep the %G and %H in the

‡ We agreed with McAvoy and Ye (1994) that the IDV(1), IDV(4) and IDV(12) + IDV(15) cases were relatively easy to handle. Since these results were not very interesting (and were essentially equivalent to those published by McAvoy and Ye), they are omitted here.

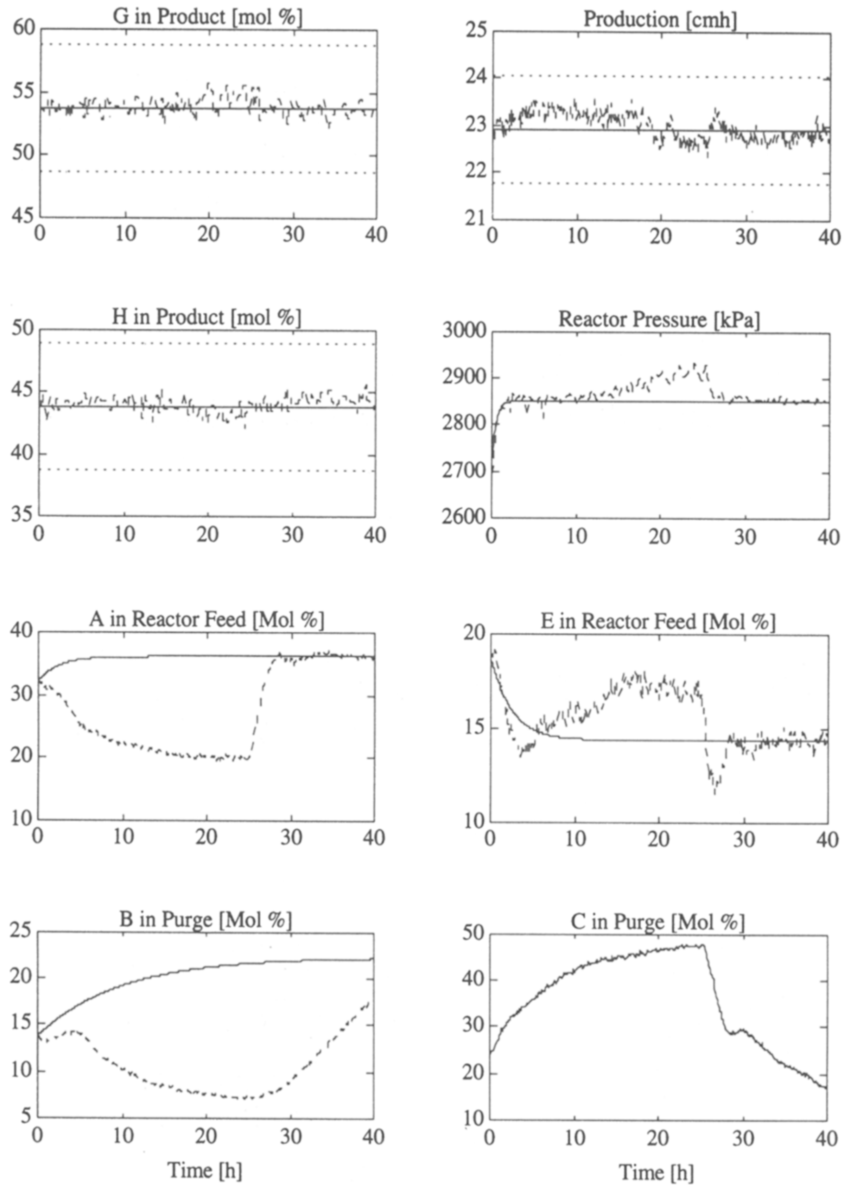

Fig. 3. NMPC controlled outputs (and selected compositions) for IDV(6)—Loss of A feed. At $t=25$, A is again available.

product well within the ±5% limits, but NMPC has lower production rate variability; the McAvoy and Ye scheme violates the ±5% limits. NMPC manipulation of the feeds is comparable to the SISO control, i.e. it does not use excessive control action.

NMPC also controls the reactor pressure more tightly. As mentioned previously, one would like to operate close to the upper shutdown limit. The McAvoy and Ye scheme would require a larger safety margin, increasing operating costs. The Desai and Rivera (1993) and Palavajjhal *et al.* (1993) multiloop schemes use the same pressure control strategy as McAvoy and Ye. It is possible that better tuning or changes in other loops would improve pressure control, but the gain changes in the response of pressure to the A feed are an inherent problem.

Loss of a feed

A more interesting case is IDV(6)$\neq$0, loss of A feed, which was discussed above in the context of cascaded flow control and constraint handling in the NMPC design. McAvoy and Ye (1994) do not present results for this disturbance, but note that: (1) an override controller is necessary, and (2) their scheme requires a reduction of 23.8% in the production rate. Others have also noted the difficulties, but have not published results. The steady-state analysis

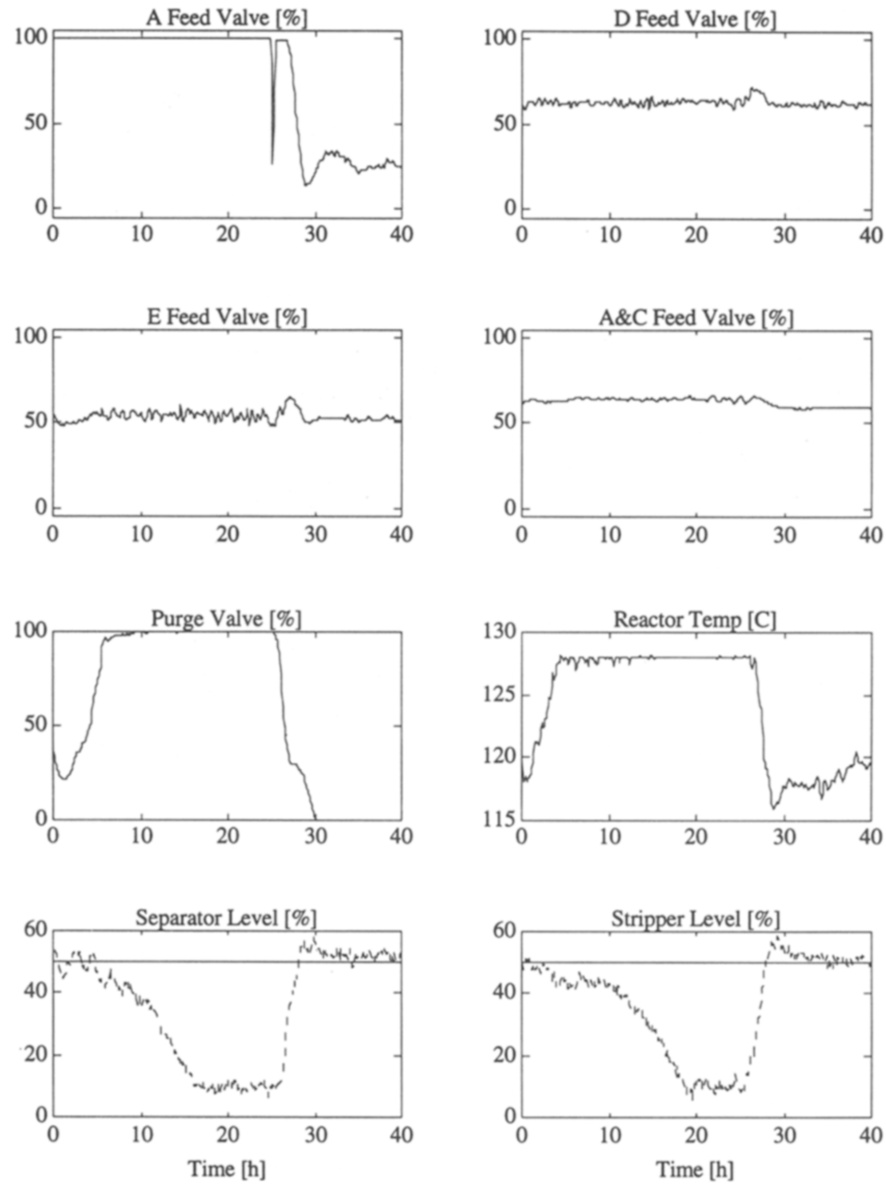

Fig. 4. NMPC manipulated variables for IDV(6)—Loss of A feed. At $t = 25$, A is again available.

of Ricker (1995) shows that full-rate production of the nominal 50/50 product mix is possible, however.

Figures 3 and 4 show how NMPC handles the loss of A feed. The disturbance occurs at $t = 0$, with the process initially at the Downs and Vogel base case condition. At the same time, we increase the reactor pressure setpoint to 2850 kPa, and begin to step the reactor level down to 60%.† Setpoints on %B in the purge, %A in the reactor feed, and %E in the reactor feed are also changed to their "optimal" values for Mode 1 (50/50) operation, as given in Ricker (1995) Table 3. Note that these *are not* optimal (or even feasible) when the A feed is lost; we assume that the operators would not know how to tune the setpoints for such a disturbance, and would probably hold them at the usual values.

At $t = 25$, the A feed returns [i.e. we set IDV(6) = 0], but no modifications are made to control system or its setpoints. Thus, the results shown in Figs 3 and 4 are for fully automatic control. Figure 3 shows that the product concentration *and* rate stay within their limits for the entire 50 h simulation. There is more product rate variability than usual, but it is acceptable. Figure 4 shows that NMPC allows the levels in the separator and stripper to drift downward as it tries to maintain production. At about $t = 17$, these hit their lower bounds, however, so from then on the reactor must keep up with the demand. Part of the NMPC strategy is to raise the reactor temperature, but this also hits its

† These setpoints represent more optimal operating conditions, as discussed previously.

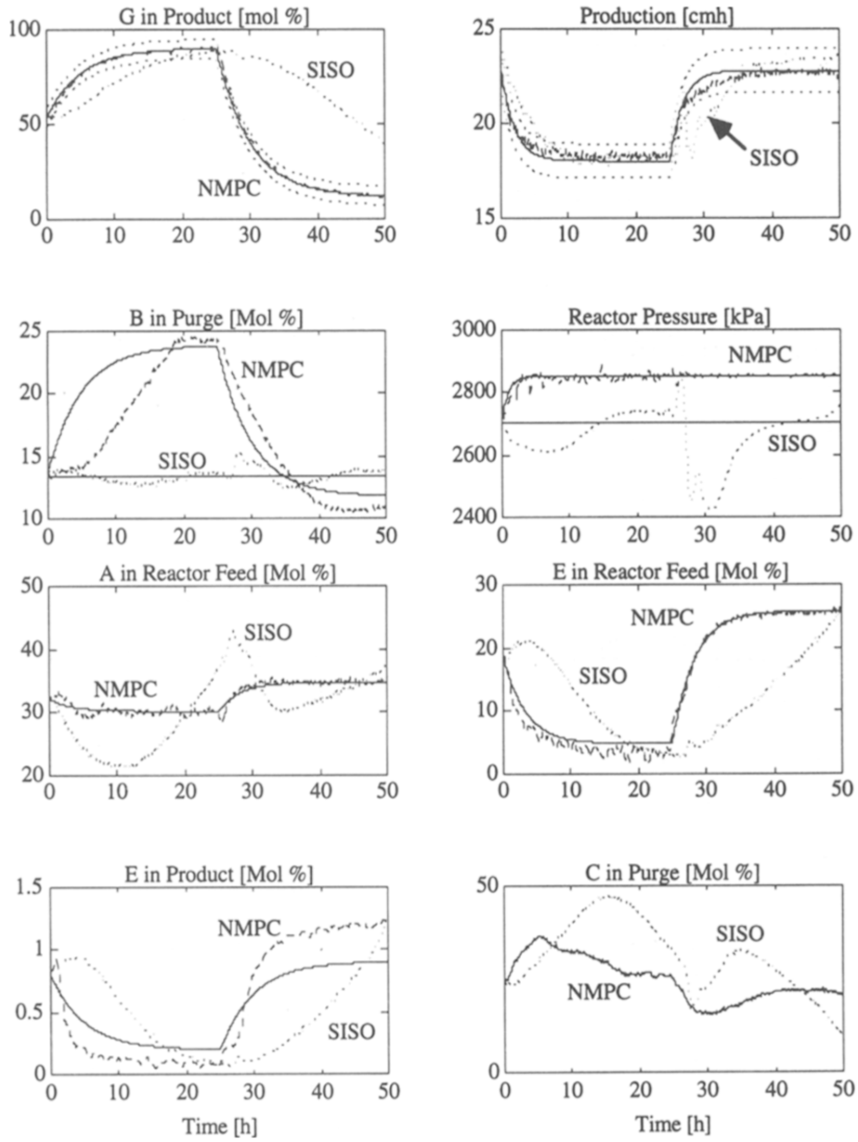

Fig. 5. NMPC vs. McAvoy and Ye (SISO) strategy for a change from 50/50 to 90/10 G/H, then from 90/10 to 10/90 G/H.

upper limit (Fig. 4), as does the purge valve. Although the A feed valve is wide open between $t=0$ and 25 (Fig. 4), the A feed flow is zero (not shown). Thus, five of the eight manipulated variables are at a constraint between $t=17$ and 25, and NMPC is still able to maintain full production at the desired composition.

Pressure control is good for the initial period (Fig. 3), but deteriorates once the liquid levels hit their lower bounds (because NMPC then needs to increase the rate of stream 4 to make more liquid, and C is already in excess as shown in Fig. 3). At about $t=22$, the pressure begins to hit $P_{r,\max}=2920$, causing the override controller to decrease the production setpoint. But the decrease is well within the ±5% limits (Fig. 3). Also, five of the eight NMPC output variables show large deviations from setpoint† (A and E in reactor feed, B in purge, and the two liquid levels), but the choice of Λ^y penalties prevents these from causing offsets in the higher priority outputs.

As discussed in the section on constraint handling, our strategy automatically resets the upper bound on the A feed rate. Thus, between $t=0$ and 25, NMPC "thinks" that only 0.2 kmol/h of the A feed is available, and it requests this maximum value. The control valve stays wide open (since even this small request cannot be satisfied). When the disturbance disappears at $t=25$, the slave flow controller rapidly shuts the valve (causing the downward spike

† The disturbance and constraints have made these setpoints infeasible.

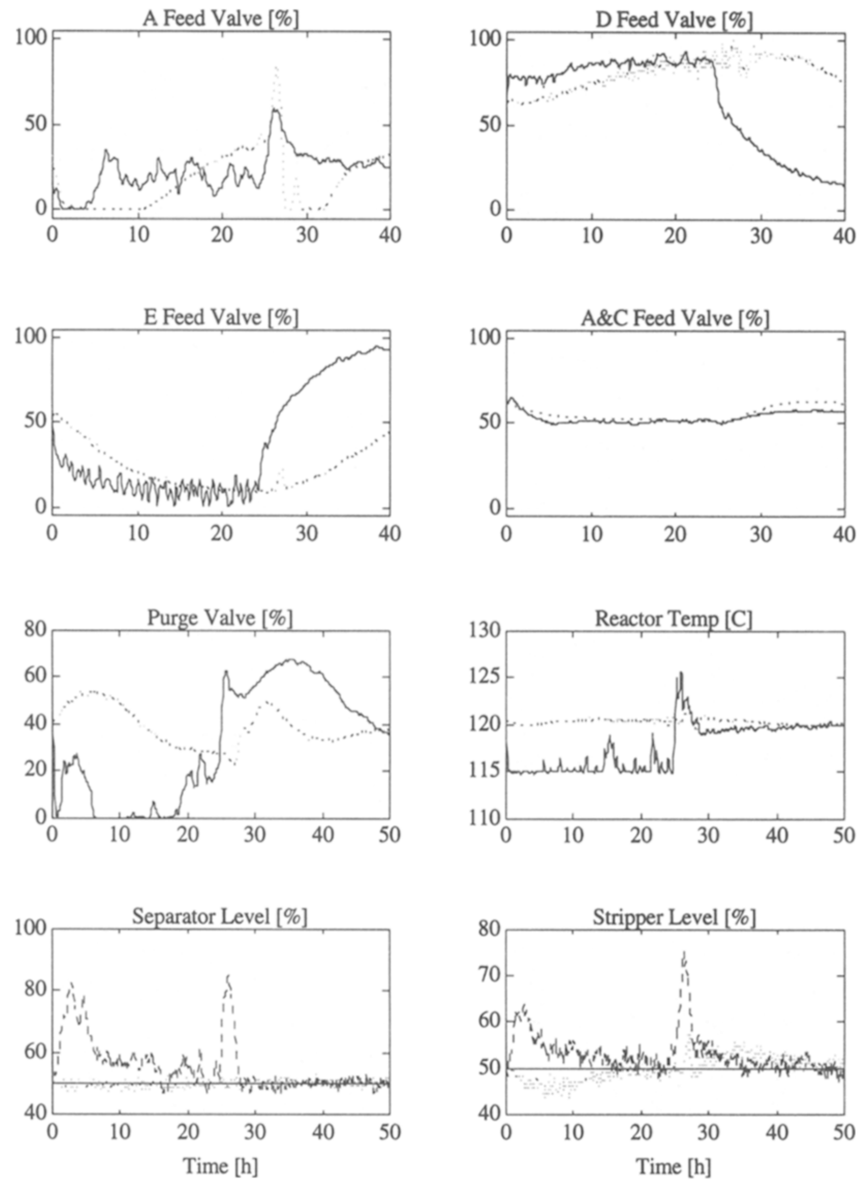

Fig. 6. NMPC vs. McAvoy and Ye (SISO) strategy for a change from 50/50 to 90/10 G/H, then from 90/10 to 10/90 G/H.

in Fig. 4). Meanwhile, however, the logic again resets the upper bound to the nominal value in Table 3. NMPC then asks for maximum flow as it attempts to get the system back to normal. These extreme moves cause small disturbances in the other feed rates, but these damp quickly as the A and E concentrations return to their "optimal" values.

Thus, we have demonstrated that the 24% loss of production cited by McAvoy and Ye (1994) can be avoided. NMPC accomplishes this by taking full advantage of the existing degrees of freedom, and operating against the active constraints, whereas the McAvoy and Ye strategy is limited by the SISO structure and the need to keep loops from saturating (or risk an unstable response).

Drift in kinetics

We also tested NMPC for drift in the reaction kinetics, disturbance IDV(13)$\neq$0. This caused no problems. The estimator detects changes in kinetics so that NMPC can make the appropriate adjustments in operating conditions. Typical performance is shown in Ricker and Lee (1995).

Setpoint changes

Downs and Vogel (1993) suggest several setpoint changes to facilitate comparisons of alternative algorithms. As we will demonstrate, NMPC easily handles setpoint changes in reactor pressure and %B in the purge that are even more extreme than those suggested by Downs and Vogel. A 15%

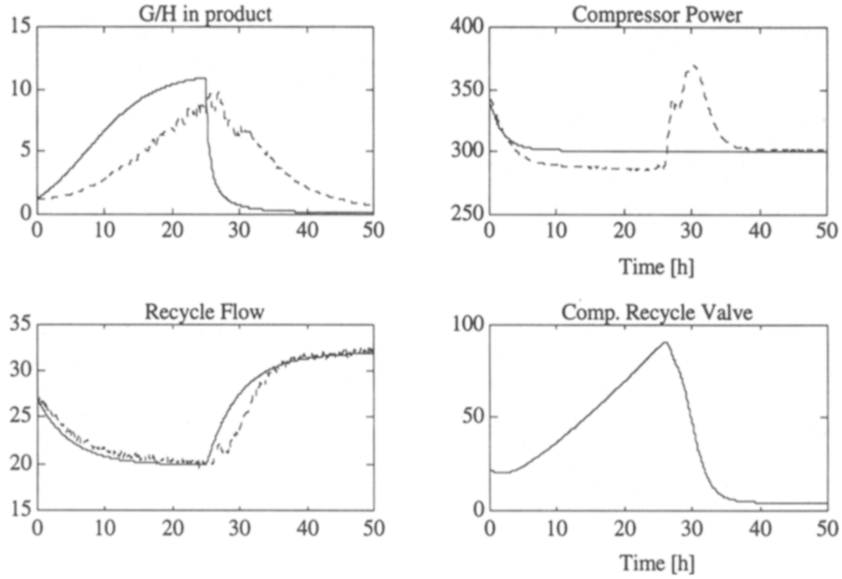

Fig. 7. Supplementary loops in McAvoy and Ye strategy for a change from 50/50 to 90/10 G/H, then from 90/10 to 10/90 G/H.

decrease in the production rate is no problem and is omitted here; in that case NMPC performance is comparable to that published by McAvoy and Ye (1994).

A more interesting test is to see how well the control system can change the product mix. Downs and Vogel suggest a change from the 50:50 base case to 40 %G, 60 %H (by weight). They also stipulate that the process should be able to operate at either 90:10 or 10:90 ratios. Figures 5–7 compare NMPC to McAvoy and Ye's method when the product composition setpoints change from the base case to 90:10, then from 90:10 to 10:90. Of course, a step change in setpoint of this magnitude is unrealistic, so we smoothed the setpoint changes for both NMPC and the McAvoy and Ye strategy (see Figs 5–7). We simultaneously varied the production rate setpoint (Fig. 5), because the base case setpoint is infeasible for 90:10 operation (Ricker, 1995). Also, for NMPC, the %B in the purge and the reactor pressure were moved to their optimal values.† This was not done for the McAvoy and Ye controller because of the poor pressure control it had exhibited in previous trials; we felt that it was more fair to leave those setpoints unchanged. We also held the reactor temperature at the base case of 120.4°C for the McAvoy and Ye controller; recall that in NMPC, it is a manipulated variable and varies automatically (see Fig. 6).

Figure 5 shows that NMPC keeps product flow and composition within bounds for the entire simulation, but the McAvoy and Ye strategy exhibits large violations. SISO control of product composition is very sluggish. This is apparently due to the tuning of the ratio controller (see Fig. 7). In separate runs we tried to make the McAvoy and Ye strategy track the %G setpoint more accurately by introducing a series of step changes in the G:H ratio setpoint, but the pressure control could not cope and eventually caused a plant shutdown. Even for the gradual changes shown in Fig. 5, the SISO pressure control saturates for extended periods (see the A feed valve in Fig. 6). NMPC pressure control is excellent. Also, although NMPC is not controlling the %E in the product, it does no worse than the SISO strategy in this respect (Fig. 5), again suggesting that manipulation of the steam flow (or equivalently, control of the stripper temperature) is unnecessary. Figure 6 shows that NMPC takes full advantage of the capacity of the separator and stripper bottoms in order to reduce product flow variations.

Figure 7 shows that in addition to the problems with the ratio control, the McAvoy and Ye loop controlling compressor power is not well-behaved.‡ We tried to choose a reasonable setpoint trajectory for this loop and the recycle-flow loop (see Fig. 7), but these specifications interact. Compressor power is also affected by the system pressure. Thus, it is

† The increase in the %B in the purge actually makes the process more difficult to control, so this was not an advantage for NMPC.

‡ The same loop appears in the schemes of Desai and Rivera (1993), and Palavajjhal *et al.* (1993).

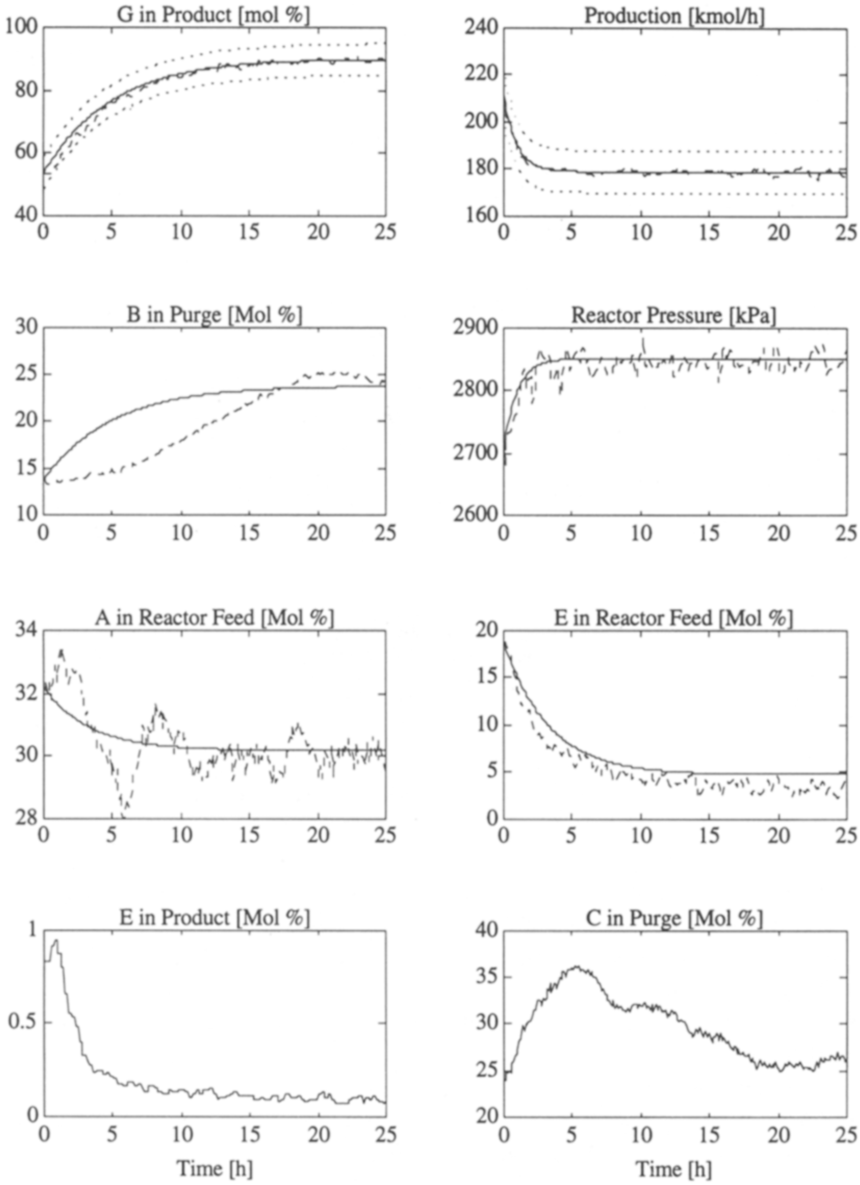

Fig. 8. NMPC for a change from 50/50 to 90/10 G/H, with five unknown disturbances, IDV(16) to IDV(20).

difficult to choose the setpoint. As shown in Fig. 7, our specification of 300 kW is a bit too high for the first part of the simulation, and the compressor recycle valve opens. In effect, it is wasting power. Then in the second phase, after a large disturbance caused by the poor pressure control, the setpoint is a bit too low, and the loop saturates in the other direction.

It is better to let the compressor power float to meet the natural demands of the process, as in our NMPC strategy. Normally, the bypass valve can be left closed, which maximizes the recycle rate. If a high recycle rate causes problems, which might be the case at low production, the operator could set this valve to a desired position and leave it there.

Unknown disturbances

Downs and Vogel (1993) provide five "unknown disturbances"—IDV(16) to IDV(20), suggesting that they should be tested during a setpoint change. Figures 8 and 9 shown NMPC performance when *all five* of these disturbances are turned on at $t=0$, and we repeat the setpoint change from the 50:50 base case to the 90:10 product composition. One may compare these results to the first 25 h of Figs 5 and 6. NMPC control of the priority variables—product rate and composition—is excellent in both cases. The only obvious signs of the disturbances are the fluctuations of the liquid levels in the separator and stripper. NMPC prevents these upsets from reaching the product.

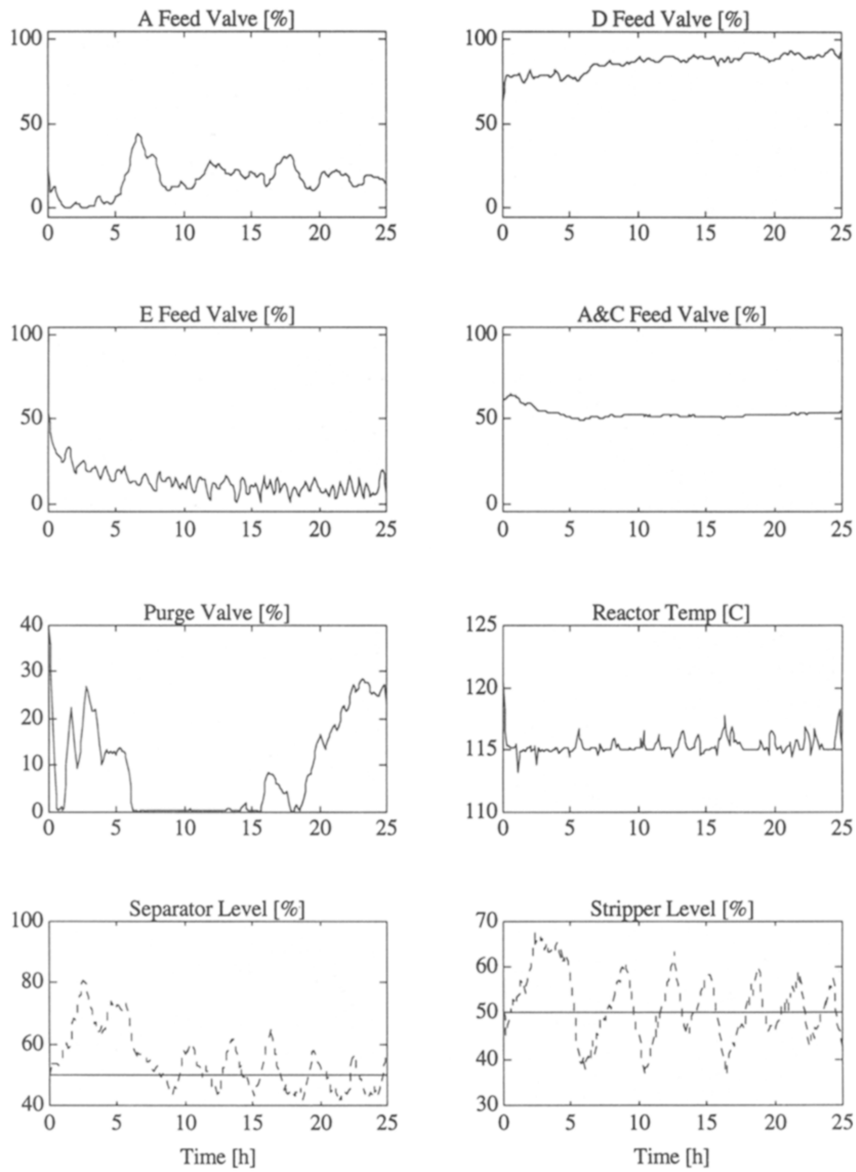

Fig. 9. NMPC for a change from 50/50 to 90/10 G/H, with five unknown disturbances, IDV(16) to IDV(20).

DISCUSSION AND CONCLUSIONS

The McAvoy and Ye (1994) multiloop strategy provides adequate performance for many scenarios, but could be improved by the elimination of the loop controlling compressor power. Manipulation of the steam flow is unnecessary, but should not degrade performance. Control of reactor pressure is adequate for the cases studied by McAvoy and Ye, but is less so when there is a large disturbance in the amount of A in the system. The use of the A feed as the manipulated variable for this loop is questioned; the gain varies with increasing A concentration, and can change sign. Other researchers (e.g. Desai and Rivera, 1993; Palavajjhala *et al.*, 1993) use the same pressure control strategy. It is possible that additional override loops would make these schemes more robust. Another possibility is the use of reactor temperature or purge rate as the manipulated variable for the pressure loop (Ricker, 1995; Banerjee and Arkun, 1994). Regardless of this choice, however, the multiloop approaches require multiple overrides to handle all the process constraints. A truly complete design of this type has yet to be published.

We have shown that NMPC (based on successive linearization and quadratic programming) improves control significantly, especially in cases involving multiple constraints and/or large setpoint changes. In the easier tests, NMPC is always better, but the difference may be too small to justify the NMPC design effort.

As in linear MPC applications, the main roadblock to NMPC was model development. The formulation of a useful nonlinear model was relatively difficult. Moreover, once the model was available, the NMPC design was not cut-and-dried. For example, the choice of manipulated and controlled variables was not obvious, and we had to test several alternatives. We were also forced to use an override loop to maintain pressure control when one of the feeds was lost.

On the positive side, model development is a good way to build process knowledge, which can be exploited *regardless* of the control strategy. As noted by Biegler and Rawlings (1991), tools for development of nonlinear mechanistic models are widely available, making such efforts a realistic option. In addition to NMPC, one could use a nonlinear, semi-theoretical model for fault detection/diagnosis, optimization, and debottlenecking.

For the version of NMPC used here, computing requirements were about the same as for integration of the full Downs and Vogel (1993) model, i.e. feasible for on-line use in a typical process computer. This is mainly due to the use of a *linear* model in the projection/control step, as suggested by Brengel and Seider (1989). We have shown that this approximation yields excellent results for the TE problem.

REFERENCES

Balchen J. G., D. Ljunngquist and S. Strand, State space model predictive control of a multistage electrometallurgical process. *Modeling, Identification and Control.* **10**, 35–51 (1988).

Banerjee, A. and Y. Arkun, Control configuration design applied to the Tennessee Eastman Plantwide Control problem. *1993 Annual AIChE Meeting*, St. Louis, Paper 148e, 11 Nov. (1993).

Bequette, B. W., Nonlinear control of chemical processes: a review. *Ind. Engng Chem. Res.* **30**, 1391–1413 (1991).

Biegler, L. T. and J. B. Rawlings, Optimization approaches to nonlinear model predictive control. In Y. Arkun and H. Ray (eds.). *Chemical Process Control—CPCIV*, pp. 543–571. AIChE/CACHE (1991).

Brengel, D. D. and W. D. Seider, Multistep nonlinear predictive controller. *Ind. Engng Chem. Res.* **28**, 1812–1822 (1989).

Desai, A. P. and D. Rivera, Controller structure selection and system identification for the Tennessee Eastman Challenge problem via intelligent process control. *1993 Annual AIChE Meeting*, St. Lous, Paper 148a, 11 Nov. (1993).

Downs, J. J. and E. F. Vogel, A plant-wide industrial process control problem. *Computers chem. Engng* **17**, 245–255 (1993).

Garcia, C. E. and A. M. Morshedi, Quadratic programming solution of dynamic matrix control (QDMC). *Proc. Am. Contr. Conf.*, San Diego (1984).

Garcia, C. E., D. M. Prett and M. Morari, Model predictive control: theory and practice—a survey. *Automatica* **25**, 335–348 (1989).

Gattu, G. and E. Zafiriou, Nonlinear quadratic dynamic matrix control with state estimation. *Ind. Engng Chem. Res.* **31**, 1096–1104 (1992).

Gelormino, M. S. and N. L. Ricker, Model-predictive control of a combined sewer system. *Int. J. Contr.* **59**, 793–816 (1994).

Grosdidier, P., B. Froisy and M. Hammann, The IDCOM-M controller. In T. J. McAvoy, Y. Arkun and E. Zafiriou (eds). *Model Based Process Control. Proceedings of the IFAC Workshop*, pp. 31–36. IFAC, Atlanta 13–14 June (1988).

Kanadibhotala, R. S. and J. B. Riggs, Nonlinear model based control of a recycle reactor process. *Computers chem. Engng* (submitted).

Lee, J. H., M. Morari and C. E. Garcia, State-space interpretation of model predictive control. *Automatica* **30**, 707–717 (1994).

Lee, J. H. and N. L. Ricker, Extended Kalman filter based nonlinear model predictive control. *Ind. Engng Chem. Res.* **33**, 1530–1541 (1994).

McAvoy, T. J. and N. Ye, Base control for the Tennessee Eastman problem. *Computers chem. Engng* **18**, 383–413 (1994).

McDonald, K. A., T. J. McAvoy and A. Tits, Optimal averaging level control. *AIChE J.* **32**, 75–86 (1986).

Morari, M. and N. L. Ricker, *Model Predictive Control Toolbox: User's Guide.* Math Works, Natick, MA (1994).

Morningred, J. D., D. A. Mellichamp and D. E. Seborg, Experimental evaluation of a multivariable adaptive nonlinear predictive controller. *J. Process Contr.* **3**, 125–134 (1993).

Muske, K. R. and J. B. Rawlings, Linear model-predictive control of unstable processes. *J. Process Contr.* **3**, 85–96 (1993).

Palavajjhala, S., R. Motard and B. Joseph, Plantwide control of the Eastman-Tenneessee problem. *1993 Annual AIChE Meeting*, St. Louis, paper 148g, 11 Nov (1993).

Patwardhan, A. A., G. T. Wright and T. F. Edgar, Nonlinear model-predictive control of distributed parameter systems. *Chem. Engng Sci.* **47**, 721–735 (1992).

Rawlings, J. B. and K. R. Muske, Stability of constrained receding horizon control. *IEEE Trans. Auto. Contr.* **38**, 1512–1516 (1993).

Ricker, N. L., The use of quadratic programming for constrained, internal model control. *Ind. Engng Chem. Process Des. Dev.* **24**, 925–936 (1985).

Ricker, N. L., Model predictive control: state of the art. Y. Arkun and H. Ray (eds.) *Chemical Process Control—CPCIV.* pp. 271–296, AIChE/CACHE (1991).

Ricker, N. L. Model predictive control of a continuous, nonlinear, two-phase reactor. *J. Process Contr.* **3**, 109–123 (1993).

Ricker, N. L., Optimal steady-state operation of the Tennessee Eastman Challenge Process. *Computers chem. Engng* **19**, 949–959 (1995).

Ricker, N. L. and J. H. Lee, Nonlinear modeling and state estimation for the Tennessee Eastman Challenge Process. *Computers chem. Engng* **19**, 983–1005 (1995).

Ricker, N. L., T. Subrahmanian and T. Sim, Case studies of model-predictive control in pulp and paper production. In T. J. McAvoy and E. Zafiriou (eds.). *Model Based Process Control—Proc. of the 1988 IFAC Workshop.* Pergamon Press, Oxford (1989).

Seborg, D.E., T. F. Edgar and D. A. Mellichamp, *Process*

Dynamics and Control, pp. 195–196. Wiley, New York (1989).

Wright, G. T. and T. F. Edgar, Nonlinear model predictive control of a fixed-bed water-gas shift reactor: an experimental study. *Computers chem. Engng* **18**, 83–102 (1994).

Ye, N., T. McAvoy, M. J. Piiiovoso and K. A. Kosanovich, Optimal averaging level control for multi-tank systems. *ADCHEM '94,* (Kyoto), IFAC, 427–432 (1994).

Zafiriou, E. and A. L. Marchal, Stability of SISO quadratic dynamic matrix control with hard output constraints. *AIChE J.* **37**, 1550–1560 (1991).

Zafiriou, E. and H. W. Chiou, Output constraint softening for SISO model predictive control. *Proc. Am. Cont. Conf.*, San Francisco, 372–376 (1993).